

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/134660/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Sierksma, Annerieke, Escott-Price, Valentina ORCID: <https://orcid.org/0000-0003-1784-5483> and De Strooper, Bart 2020. Translating genetic risk of Alzheimer's disease into mechanistic insight and drug targets. *Science* 370 (6512) , pp. 61-66. 10.1126/science.abb8575 file

Publishers page: <http://doi.org/10.1126/science.abb8575>  
<<http://doi.org/10.1126/science.abb8575>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



# Translating genetic risk of Alzheimer's disease into mechanistic insight and drug targets

Annerieke Sierksma<sup>1,2</sup>, Valentina Escott-Price<sup>3,4,\*</sup>, Bart De Strooper<sup>1,2,5,\*</sup>

<sup>1</sup> VIB Center for Brain & Disease Research, Leuven, Belgium

<sup>2</sup> Laboratory for the Research of Neurodegenerative Diseases, Department of Neurosciences, Leuven Brain Institute (LBI), KU Leuven (University of Leuven), Leuven, Belgium

<sup>3</sup> Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff

<sup>4</sup> UK Dementia Research Institute, Cardiff University, Cardiff, UK

<sup>5</sup> UK Dementia Research Institute, University College London, London, UK

\*to whom correspondence can be addressed

## Abstract

In order to provide better prevention and treatment we need to understand the environmental and genetic risk of Alzheimer's Disease (AD). However, the definition of AD has been confounded with dementia in many studies. Thus, over interpretation of genetic findings with regard to mechanisms and drug targets may explain in part controversies in the field. Here, we analyze the different forms of genetic risk of AD and how these can be used to model disease. We stress the importance of studying gene variants in the right cell types and in the right pathological context. The lack of mechanistic understanding of genetic variation has become the major bottle neck in the search for novel drug targets for AD.

## Introduction

The number of people worldwide suffering from dementia is now already slightly exceeding the number of people with cancer, and is poised to increase even more over the next decades. Dementia is, however, a container term for the end symptoms of a wide variety of brain diseases, including Alzheimer's Disease (AD). AD is a slowly progressing disorder characterized by specific protein accumulations in the brain. Clinical dementia manifests only late, which confounds many case-control studies using this criterion as a proxy for AD. Prodromal AD patients become excluded and not yet cognitively altered AD cases are mixed with controls. About 30% of clinically diagnosed patients have no neuropathological or biomarker characteristics of AD (1) and 56% of cases defined as AD present with common comorbidities such as Lewy body disease, vascular pathology or hippocampal sclerosis (2). Unfortunately, the advice of the National Institute of Aging and the Alzheimer Association Research Framework to define AD as a biological construct (3), is not yet widely adopted.

A recent comprehensive overview estimated that 35% of life time risk of dementia is modifiable, including factors such as education, vascular aspects, hearing loss, social deprivation, etc. (4). The Framingham Heart Study confirms the modifiable nature of dementia risk, with decreasing incidence of dementia over the last decades (5). However, while this trend was very significant for dementia overall and for vascular dementia in particular, it was not significant for AD alone (5). Given the high heritability of AD (6, 7), studying genetic risk seems a more fruitful way forward to identify molecular mechanisms of disease. The central question is whether there is one central route to AD, and therefore one 'type' of AD, or whether various pathogenic mechanisms exist that converge on the defining amyloid plaque and tangle pathology.

## The heritability of AD

Heritability, formally defined as the proportion of phenotypic variance that is due to genetic factors, can be used as a population-based measure for the risk of disease (see glossary in box 1).

Importantly, the inheritance of genetic risk variants does not necessarily imply disease and not all individuals with AD carry the same risk variants.

The best studied risk (or better causal) genetic variants in AD are the fully penetrant mutations in the genes encoding Amyloid Precursor protein (*APP*) and presenilin 1 and 2 (*PSEN1&2*). They affect the processing of the amyloid- $\beta$  peptide ( $A\beta$ ) indicating that  $A\beta$  aggregation is an upstream event in the pathogenesis of AD (8). These mutations were identified in families with a Mendelian, dominantly inherited form of AD (8) which clinically manifests as early onset dementia (onset before <65 years). Estimates for the heritability of early onset AD are very high, ranging between 0.92 and 1 (7). Even in this smaller group (<10% of total AD patients), *APP* and *PSEN1&2* mutations explain only about 10% of these early onset cases (7). The remaining heritability is explained by *APP* duplications, by an increasing number of rare variants in genes encoding e.g. the Sortilin-related receptor (*SORL1*), Triggering receptor expressed on myeloid cells 2 (*TREM2*) and ATP Binding Cassette Subfamily A Member 7 (*ABCA7*) (9), and finally by not yet identified, but likely recessive, mutations (7). An example of a recessive AD mutation is A673V in *APP* (10).

In the large group of patients in which dementia manifests after age 65, the heredity is also large, estimated between 0.58 and 0.78 with rather large 95% confidential intervals [0.19-0.87] and [0.67-0.88] respectively (6). This is high compared to other late onset diseases. The genetic architecture underlying AD >65 years is far from fully charted (see Fig.1). Apolipoprotein  $\epsilon$ 4 allele (*APOE4*) is the only common high risk genetic variant (odds ratio (OR)=3.32) (11, 12). Genome wide association studies (GWAS) have further identified many common genetic variants with low risk (OR=1.1-1.2) of which 40 have genome wide significance (12–14). Exome chip analyses have additionally yielded rare variants in the very same genes, i.e. *SORL1*, *TREM2*, and *ABCA7*, that increase strongly risk of early onset AD (9, 15). Variants that protect have also been discovered: the *APOE2* allele (OR=0.6) (16), rare mutations in phosphatidylinositol specific phospholipase C-gamma 2 (*PLC $\gamma$ 2*) (OR=0.68)(15) and the Icelandic mutation A673T in *APP* (OR=0.19) (10). The genetics of early and late

onset disease suggest that AD should be considered a continuum. As indicated by the broad confidential intervals, the AD heritability estimates remain imprecise, as with many polygenetic disorders (17). Increasing efforts to create larger datasets for GWAS studies, to directly sequence full genomes and to develop novel data analyses methodologies are under way to tackle the “missing heritability” in AD.

## **From heritability to mechanisms of disease**

Translating genetic information into disease mechanisms is anything but trivial. It took twenty years to understand that AD causing mutations destabilize presenilin leading to premature release of long A $\beta$  peptides (18). Similar efforts to understand how rare mutations in the open reading frame of *TREM2*, *PLC $\gamma$ 2*, *SORL1* and *ABCA7* affect protein function will be needed. In addition, most available genetic information in AD remains imprecise. The causal variant is known for only 40% of the identified GWAS loci (14), the effect of these variants is only known in a minority of cases, and literally thousands of these variants contribute to the heritability of the phenotype.

The question is what the core genes are, i.e. which genes execute a direct effect on the disease process. Unfortunately, more than 70% of variants determining phenotypic variation are in “peripheral” genes. Such genes have only indirect effects on expression or posttranslational modification of core gene products and as such are not very informative for the molecular mechanisms driving the phenotype (19). The individual ‘trans’ effects of these peripheral genes are small (19) but since there are many, they underlie a large part of the heritability (20, 21). Even using the p-value of  $p < 5 \times 10^{-8}$  for genome wide significance to prioritize gene loci (which comprise now 40 loci in AD (14)) does not provide certainty of finding “core disease pathway genes”(19). The frustrating conclusion is that the bulk of the heredity in AD likely only indirectly points to key biological pathways of disease. One group of peripheral genes, i.e. master regulator genes, is nevertheless of particular interest. These genes, encoding for example transcription factors, chromatin modifiers, regulatory RNA or enzymes,

regulate the expression or function of several disease core genes. For example the AD risk locus *Spi-1* proto-oncogene (*Spi1*) codes for the transcription factor Pu.1 which regulates many microglia genes, pointing to a role for inflammation in AD (22). Such master regulators are usually under strong evolutionary constraint, and so not easily detected in GWAS (19).

One could try to investigate how peripheral genes affect the expression of core genes. A prerequisite is to understand in which cells these peripheral genes exert their effect and hence single cell analyses of gene expression in brain cells is crucial (23). Such trans-expression quantitative trait loci (trans-eQTL) mapping, however, needs huge data sets and is only readily available for peripheral blood cells. Another possibility is to focus on gene variants with large effects on heredity. ApoE4 is the only example in AD. Finally, one could ignore the quantitative contributions of genes to heredity and focus on rare variants, which are likely more central to the disease mechanisms because of their large effect sizes. A potentially fruitful avenue of research is to investigate how the common variants that define heredity regulate these rare variant genes.

## **The complexity of the APOE locus**

The 3 major isoforms of ApoE ( $\epsilon 2$ ,  $\epsilon 3$  and  $\epsilon 4$ ) are defined by two single nucleotide polymorphisms or SNPs (rs429358 and rs7412) within exon 4 of the gene (11). The  $\epsilon 4$  allele (frequency 0.14 in the Caucasian population) provides a 3-fold increased risk of AD which raises to 14-fold in  $\epsilon 4$  homozygotes (11, 16). Conversely, the  $\epsilon 2$  allele (frequency 0.08) confers a 1.7-fold decreased risk. This risk is more pronounced in women than in men and is strongly depending on ethnic background, i.e. the  $\epsilon 4$  effect is much smaller in the African-American and Hispanic population (16). This illustrates the importance of multi-ethnic genetic studies when studying the heritability of AD.

The APOE locus is highly complex, spanning almost 2MB and covering over 70 genes. Despite being in low linkage disequilibrium with the APOE SNPs for  $\epsilon 2$  and  $\epsilon 4$ , there are many other SNPs in this large locus showing significant association to AD. This might point to other AD risk genes in this

locus. Several of these SNPs, however, likely affect expression of APOE. Understanding this will be of tremendous value as it would clarify whether and under which conditions up- or down regulation of this multifunctional protein could affect the risk of AD.

Under physiological conditions, APOE is mainly expressed by astrocytes, but microglia exposed to A $\beta$  plaques highly upregulate APOE. It will be critical to unravel how microglial function is affected by different APOE isoforms and how this contributes to disease. Knock-out of the gene eliminates the AD induced inflammatory response in mice (24).

While very relevant to AD, the role of APOE in brain inflammation remains badly understood. APOE obviously plays a crucial role in cholesterol transport and lipid homeostasis, but also in A $\beta$ -aggregation, -clearance, and -cellular uptake, and affects, via less well understood molecular pathways, synapse number and function, blood-brain barrier integrity, and TAU-mediated neurodegeneration (16, 24, 25). It is important to decipher which roles of APOE are directly relevant to AD as the variety of functional effects of APOE deficiency in different cell types and in different tissues suggest that the APOE gene is a master peripheral regulator in the disease. Not all affected pathways are necessarily relevant to AD. Directly modulating APOE to protect against AD is likely to have a variety of effects and the outcome of such treatments will need careful monitoring.

### **Causal, high risk and protective variants are involved in APP processing and in microglial function**

Evidence of genotype-phenotype dose-responses in an allelic series strongly argues for a core gene function. Such gene-dosage effects are observed with genes involved in A $\beta$ -generation. Next to the fully penetrant *APP* and *PSEN* mutations, *APP* gene duplications and triplications, including Down syndrome, cause AD (8, 10). A recessive (E673V) and a protective (E673T) allele (10) affect the propensity of A $\beta$  to aggregate. E673T also lowers  $\beta$ -secretase processing of APP (26). A common allele (rs2154481) in the *APP* locus lowers risk (OR=0.95) although, counter intuitively, slightly increases APP

expression (14). Finally, variants in the gene locus of the  $\alpha$ -secretases, *ADAM17* and *ADAM10* (27) all demonstrate that APP itself and the enzymes processing it to A $\beta$  carry risk and even cause AD.

*SORL1* provides another example of an allelic series with increasing risk of AD. *SORL1* encodes the sorting-related receptor with A-type repeats SorLA (a.k.a. LR11) involved in retromer-related endosomal traffic. SorLA contains functional domains that can bind monomeric A $\beta$  or APP. Several of the deleterious variants affect those domains (28). SorLA lowers A $\beta$  production by redirecting APP to the cell membrane and trans-Golgi network (TGN) and A $\beta$  to lysosomes in neurons (see Fig.2)(28). Interestingly, *SORL1* expression is twentyfold higher in human than in mouse microglia, warranting for further characterization of the impact of *SORL1* deficiency on microglia functions (29). The different *SORL1* variants aggregate into categories with increasing risk burden (30) from OR=1.21 for missense variants up to OR=16.73 for protein-truncating variants (9). These ORs are comparable to heterozygous (OR=3.2) and homozygous (OR=14.9) APOE- $\epsilon$ 4 carriers. Variants are present in 2% of AD, compared to <1% for APP and PSEN1 mutations (30).

Common and rare *ABCA7* variants provide a third allelic series. *ABCA7* promotes the efflux of phospholipids out of cells. Protein-truncating (OR=2.6) and missense mutations (OR=1.8) are associated with AD (31). In addition, a tandem repeat in intron 18, ranging from 300 bps to more than 10kb, provides relative high risk of AD (OR=4.5). It remains unclear how loss of function of *ABCA7* increases risk of AD, although in mice it causes higher A $\beta$  plaque burden related to impaired A $\beta$  phagocytosis in macrophages and microglia (32). Loss of *ABCA7*, because of its role in lipid transport, might have broad effects on cell physiology (see Fig.2). The relative low OR suggests that it is not directly causally involved and its broad function suggest it acting as a master regulator peripheral gene in AD.

Next to core and master regulator genes affecting A $\beta$ , strong genetic evidence imply microglia in AD. Many common variants associated to risk of AD occur in genes that are expressed in microglia



(see table 1 in (33)). Rare missense mutations in the open reading frames of *TREM2*, *PLCγ2* and *ABI3* (Abelson interactor family protein 3) which are genes mainly or exclusively expressed in myeloid cells, also point into that direction (15). Many of the risk genes of AD become upregulated in microglia when exposed to Aβ but less so to TAU pathology (34). Thus, a large part of the genetic risk of AD, as opposed to genetic cause of AD, seems to converge into the microglial response to amyloid plaques.

One of the best studied genes in this series is *TREM2*. *TREM2* is a receptor for anionic ligands including phospholipids, lipopolysaccharide (LPS) and DNA (35). In mouse models of AD *Trem2* is required for the transit of microglia from homeostatic to activated cell states in response to amyloid plaques. Interestingly these microglia strongly upregulate ApoE expression (36). *Trem2* deficiency leads to more diffuse plaques with greater neuritic damage and less recruitment of microglia to amyloid plaques (36). The rare R47H and more common R62H variants of *TREM2* (15) alter its stability, affect phagocytic capacity and impair *TREM2* affinity for APOE, clusterin (ApoJ), low-density lipoproteins (LDL) and Aβ (33). The impact of other more common variants on *TREM2* function remains unclear. Since the R47H and R62H mutations cause partial loss of function of *TREM2*, and since *Trem2* deficiency seems to aggravate amyloid plaque pathology in mice, most drug development efforts are focused on enhancing *TREM2* function (37). However, enhancing microglia activity might be a two edged sword with opposite effects on Aβ and TAU pathology (24). It also remains a big question to what extent observations in mice can be extrapolated to the human pathophysiology as the cellular reactions around amyloid plaques are much more complex in human than in the available mouse models (see for instance (38, 39)). The fact that several additional AD associated variants have been observed in genes that act downstream of *TREM2* nevertheless underlines the importance of *TREM2*-signaling in AD (see Fig.2).

One example is the rare protective P522R variant in the *PLCγ2* gene with moderate effect size (OR=0.57-0.68) (15). This mutation increases the activity of the microglial signaling enzyme phospholipase C gamma 2 downstream of *TREM2* (see Fig.2). The variant is overrepresented in a

cohort of cognitively healthy centenarians and, anecdotally, provides full protection to *APOE-ε4* in a more than hundred years old homozygous carrier (40). PLCγ2 becomes phosphorylated upon stimulation and affects phagocytosis, migration, chemokine and cytokine release (41). Structurally, the P522R variant modifies an autoinhibitory domain of PLCγ2 leading to greater PIP<sub>2</sub> conversion and increased cellular calcium release (41). The P522R variant enhances Aβ endocytosis, suggesting that this protective variant may facilitate microglial clearance of Aβ.

In conclusion, the genetics of AD provide strong evidence for a major pathway centered on Aβ generation, aggregation and clearance that operates in early and late onset disease. The genetics also strongly imply microglia responses to amyloid plaques in AD. Assuming that these responses are directed by the genetic risk profile of the patient, one would predict that some patients are protected from the damage caused by amyloid plaques because of their excellent microglia (34). Major questions for the field are what aspects of the microglia response on amyloid plaques are beneficial or detrimental, how genetic risk affects this balance, and whether this contributes to TAU pathology. Drug development will have to move cautiously taking into account this fine Yin- and Yang of the cellular response in AD (23).

## **Leveraging polygenic risk**

A large proportion of the genetic risk of AD is explained by common variation in the genome and is captured via single nucleotide polymorphisms (SNP) in GWAS (21). Such single variants on their own do not predict an individual's risk of AD but can be combined in a polygenic risk score (PRS). PRS is a “genetic score” defined as the sum of the number of SNP risk alleles that an individual carries, weighted by their contribution to the disease risk (effect size).

Most investigators currently use a partial AD PRS calculated with the lead SNP in the 40 canonic GWAS genome loci mentioned before (12–14, 42). However, a more complete PRS calculation includes the thousands of other SNPs in loci that are associated with risk of AD but did not reach the threshold

for genome wide significant association ( $p < 5 \times 10^{-8}$ ). Such calculation improves strongly the prediction accuracy of AD, something also observed with psychiatric and other complex disorders (43). In fact, the prediction accuracy of AD using the complete PRS is high, with area under the receiver-operator curve (AUC) of 75% in clinical and 84% in pathologically confirmed samples (21, 44).

Using only the canonic GWAS loci biases the score to the effect of the APOE region (21). If all genetic risk of AD is used as proposed for the complete PRS, the bulk of associated SNPs of small effect sizes will eventually outperform the effect size of the APOE locus alone. Accordingly, the predictive accuracy of complete PRS in pathologically confirmed E3 homozygotes is high, with  $AUC > 80$  (45). To date, the PRS approach has mostly been assessed in European populations due to lack of multiethnic GWAS data.

The field is currently struggling to translate the concept of PRS into meaningful functional hypotheses. An interesting recent development is to include only SNPs associated with genes from putative disease-specific pathways, for instance APP metabolism, lipid metabolism, endocytosis, to generate pathway specific PRS (27). However, the definition of these disease pathways is based mostly on the different functional categories defined by Gene Ontology (46). This is problematic (46) as there is little expert scrutiny, inclusion thresholds are low and almost all AD genes are implicated in more than one pathway (12, 47, 48). It turns out that the AD predictability using such categories is low (47). Finally, it is important to mention that the PRS is currently designed as a linear combination of SNP effect sizes without accounting for non-linear effects, also known as epistasis or SNPxSNP interaction. Biologically, it is very unlikely that genetic risk of AD is the simple additive sum of the individual SNP risks.

## **From polygenic risk to mechanisms of disease and drug targets: cellular state and disease context matter**

Drugs developed against targets supported by genetic evidence have a better chance to become approved (49). However, in the AD field the causal SNPs are in many cases unknown or

assigned to the wrong gene. There are still large gaps in understanding how SNPs affect the functional genomic architecture. At the other side, information on the effects of drugs on eQTL are usually not in the public domain, making it difficult to link experimental drugs to candidate targets. Overall, the single most important limiting factor in the translation of knowledge from genetics to drugs is, however, the lack of good models for AD (see Fig.3).

Assessing the functional impact of non-coding risk variants is challenging and starts with the question whether a particular SNP is functional or is only in linkage disequilibrium with the real functional SNP. Risk mechanisms will only manifest in disease-relevant conditions and therefore cell type and experimental context really matters when analyzing the functional consequences of SNPs (23). Finally, SNPs are frequently assigned to genes that are most close in the linear DNA sequence representation of the genome. However, chromatin has a complex 3D structure and enhancers or suppressors can exert their effects on the expression of genes that are remote from their location (50). Recent work has indeed shown that many causal variants affect enhancers which are highly specific to brain region, cell-type, and cell state (22, 50, 51). It was noted that mainly myeloid and microglial enhancer regions, and not the promoter regions are significantly enriched for AD-associated variants (50–52). An elegant knock-out experiment underscored this conclusion. Nott et al. deleted in human induced pluripotent stem cells (iPSCs) the *BIN1* (Bridging Integrator-1 and Amphiphysin-2) enhancer which carries one of the higher AD risk variants, rs6733839 (OR=1.2)(12, 50). When they differentiated these cells into microglia, astroglia, and neurons, expression of BIN1 was only affected in microglia (50).

The AD field really struggles to generate good models that reproduce all features of disease (see Fig.3). Double and even triple transgenic mice overexpressing human TAU, PSEN and APP, all with FAD or FTD mutations, are needed to obtain amyloid plaques and tangles and it remains a tantalizing question to what extent cellular phenotypes induced in these mice mimic the situation in human. Sixty-

five million years of evolution divergence cannot be ignored when modeling a human polygenic disease. In order to research human-specific cell biology, research on human iPSCs has taken a flight, including *in vitro* 3D (53) and organoid cultures (54). All are promising, but each approach comes with its own limitations (see Fig.3). For instance, the 3D *in vitro* cultures provide a very artificial conformation to grow the cells and uses high overexpression of the APP gene in the neurons to obtain AD phenotypes (53). The human organoid culture is promising but their usefulness to study non-developmental disorders remains debated (54).

The xenograft or chimeric, mouse model approach, where human iPSC-derived brain cells are transplanted into the mouse brain (55–57) provides an interesting alternative combining several advantages. The rodent brain functions as a superior “physiological” 3D matrix for human cells compared to other more artificial environments. Human neurons (55), microglia (56, 57) and astroglia have been grown in rodent brains for over one year and reproduce many human features. While the rodent brain background and the immune suppression are confounders in these experiments, microglia cells, even after exposure to a cell culture environment, fully regain their identity when returned to the CNS and transcriptionally closely resemble freshly isolated human microglia from surgical samples (57).

In theory, human iPSCs and their derived models can be used to functionally evaluate the impact of PRS-defined risk in different cell types and AD-relevant contexts. Obviously, the genomic variants captured from different patients will be different, but, because the pathological phenotype of AD patients is very similar, it is assumed that the cellular pathology converges and that shared pathways leading to disease may be identified. Once a critical mass of PRS-defined iPSCs will have been analysed, one can also envision eQTL and regulatory landscape analyses to define how specific AD-associated variants may exert their effects. This can subsequently refine the list of SNPs, including only core variants driving AD pathogenesis. Ultimately, such functional insights will lead to better and more

relevant PRS that will be used for diagnostics, stratification of patients for clinical trials, and personalized medicine based on genetic profile.

## Conclusions

The genetic component in AD risk is surprisingly large for a late onset disorder. Tremendous progress has been made to map this genetic landscape, but now it becomes critically dependant on a better definition of AD and the underlying mechanisms of disease. “More” cases is never going to replace “quality” of cases and deeper clinical phenotyping and biomarkers are needed to better interpret the role of genetic variation in specific aspects of the AD phenotype.

It is crucial, while working further along those lines, but also from therapeutic development perspective, to take into account the long preclinical phase of AD (23). At the functional level we need to get away from the classical molecular biology paradigms of gene-function-drug target. Gene variants affect gene function in specific genetic backgrounds (mice are not humans), in specific cell types, in specific cell states, and in specific stages of the disease. In silico predictions and simple cell biological experiments, although tempting because of the high throughput, can be very misleading and can bring a whole drug development campaign in jeopardy. Finding drugs for a complicated multifactorial disease like AD requires deep knowledge of the mechanisms that are targeted. The full mapping of the cellular phase of AD is now a priority for the field (23).

One should, however, acknowledge the tremendous progress made in AD research. We can now further build on the many hints coming from genetic work over the last decade to generate the more sophisticated models that will better represent specific mechanisms underlying AD. This novel thinking will open many opportunities for drug development, while better stratification of patients will accelerate the road from concept to clinic.

## References and Notes

1. T. G. Beach, S. E. Monsell, L. E. Phillips, W. Kukull, Accuracy of the Clinical Diagnosis of Alzheimer Disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *J. Neuropathol. Exp. Neurol.* **71**, 266–273 (2012).
2. G. D. Rabinovici *et al.*, Multiple comorbid neuropathologies in the setting of Alzheimer’s disease neuropathology and implications for drug development. *Alzheimer’s Dement. Transl. Res. Clin. Interv.* **3** (2017), pp. 83–91.
3. C. R. Jack *et al.*, NIA-AA Research Framework: Toward a biological definition of Alzheimer’s disease. *Alzheimer’s Dement.* **14** (2018), pp. 535–562.
4. G. Livingston *et al.*, Dementia prevention, intervention, and care. *Lancet.* **390**, 2673–2734 (2017).
5. C. L. Satizabal *et al.*, Incidence of dementia over three decades in the Framingham heart study. *N. Engl. J. Med.* **374**, 523–532 (2016).
6. M. Gatz *et al.*, Role of genes and environments for explaining Alzheimer’s disease. *Arch. Gen. Psychiatry.* **63**, 168–74 (2006).
7. T. S. Wingo, J. J. Lah, A. I. Levey, D. J. Cutler, Autosomal recessive causes likely in early-onset Alzheimer disease. *Arch. Neurol.* **69**, 59–64 (2012).
8. R. E. Tanzi, L. Bertram, Twenty years of the Alzheimer’s disease amyloid hypothesis: A genetic perspective. *Cell.* **120**, 545–555 (2005).
9. C. Bellenguez *et al.*, *Neurobiol. Aging*, in press, doi:10.1016/j.neurobiolaging.2017.07.001.
10. J. TCW, A. M. Goate, Genetics of  $\beta$ -Amyloid Precursor Protein in Alzheimer’s Disease. *Cold Spring Harb. Perspect. Med.* **7** (2017), p. a024539.
11. E. H. Corder *et al.*, Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer’s disease in late onset families. *Science (80-. ).* **261**, 921–923 (1993).
12. B. W. Kunkle *et al.*, Genetic meta-analysis of diagnosed Alzheimer’s disease identifies new risk

- loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
13. I. E. Jansen *et al.*, Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
  14. I. de Rojas *et al.*, Common variants in Alzheimer's disease: Novel association of six genetic variants with AD and risk stratification by polygenic risk scores. *medRxiv*, 19012021 (2020).
  15. R. Sims *et al.*, Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat. Genet.* **49**, 1373–1384 (2017).
  16. M. E. Belloy, V. Napolioni, M. D. Greicius, Review A Quarter Century of APOE and Alzheimer's Disease: Progress to Date and the Path Forward. *Neuron.* **101**, 820–838 (2019).
  17. A. J. Mayhew, D. Meyre, Assessing the Heritability of Complex Traits in Humans: Methodological Challenges and Opportunities. *Curr. Genomics.* **18**, 332–340 (2017).
  18. M. Szaruga *et al.*, Alzheimer's-Causing Mutations Shift A $\beta$  Length by Destabilizing  $\gamma$ -Secretase-A $\beta$ n Interactions. *Cell.* **170**, 443-456.e14 (2017).
  19. X. Liu, Y. I. Li, J. K. Pritchard, Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell.* **177**, 1022–1034 (2019).
  20. R. S. Desikan *et al.*, Genetic assessment of age-associated Alzheimer disease risk: Development and validation of a polygenic hazard score. *PLOS Med.* **14**, e1002258 (2017).
  21. V. Escott-Price *et al.*, Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain.* **138**, 3673–84 (2015).
  22. K. Huang *et al.*, A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nat. Neurosci.* **20**, 1052–1061 (2017).
  23. B. De Strooper, E. Karran, The Cellular Phase of Alzheimer's Disease. *Cell.* **164**, 603–615 (2016).
  24. J. M. Long, D. M. Holtzman, Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. *Cell.* **179** (2019), pp. 312–339.
  25. A. Montagne *et al.*, APOE4 leads to blood–brain barrier dysfunction predicting cognitive



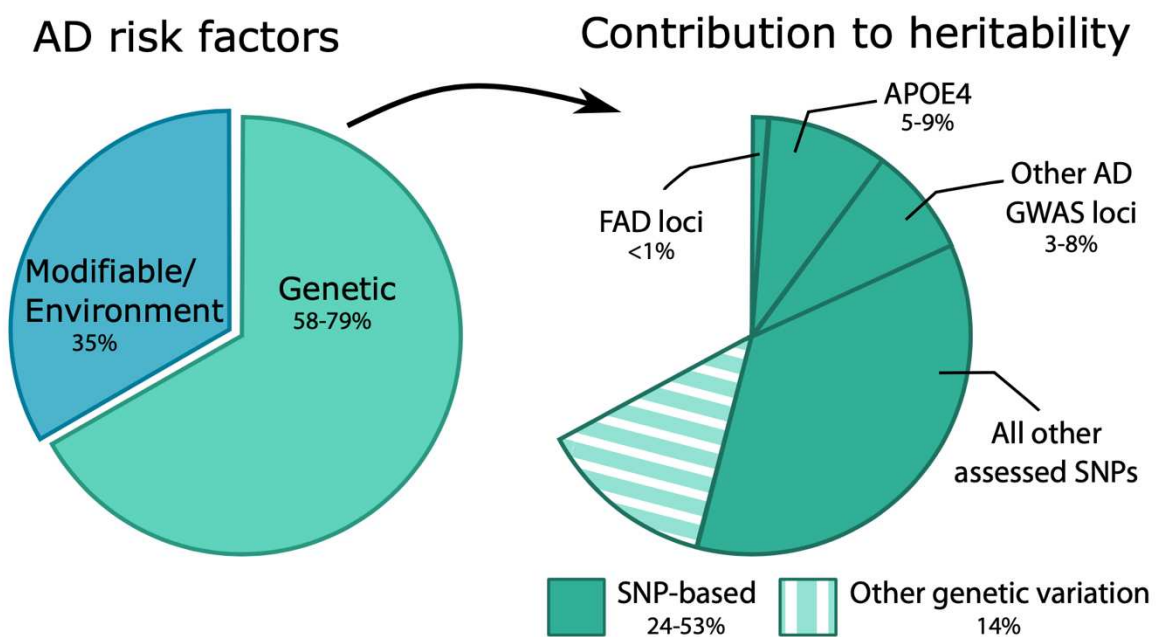
- decline. *Nature*. **581** (2020), doi:10.1038/s41586-020-2247-3.
26. I. Benilova *et al.*, The Alzheimer disease protective mutation A2T modulates kinetic and thermodynamic properties of amyloid- $\beta$  (A $\beta$ ) aggregation. *J. Biol. Chem.* **289**, 30977–30989 (2014).
  27. R. Sims, M. Hill, J. Williams, The multiplex model of the genetics of Alzheimer's disease. *Nat. Neurosci.* **23**, 311–322 (2020).
  28. D. Campion, C. Charbonnier, G. Nicolas, SORL1 genetic variants and Alzheimer disease risk: a literature review and meta-analysis of sequencing data. *Acta Neuropathol.* **138** (2019), pp. 173–186.
  29. D. Gosselin *et al.*, An environment-dependent transcriptional network specifies human microglia identity. *Science (80-. ).* **356**, 1248–1259 (2017).
  30. H. Holstege *et al.*, Characterization of pathogenic SORL1 genetic variants for association with Alzheimer's disease: A clinical interpretation strategy. *Eur. J. Hum. Genet.* **25**, 973–981 (2017).
  31. A. De Roeck, C. Van Broeckhoven, K. Sleegers, The role of ABCA7 in Alzheimer's disease: evidence from genomics, transcriptomics and methylomics. *Acta Neuropathol.* **138** (2019), pp. 201–220.
  32. T. Aikawa *et al.*, ABCA7 haplodeficiency disturbs microglial immune responses in the mouse brain. *PNAS.* **116**, 23790–23796 (2019).
  33. D. V Hansen, J. E. Hanson, M. Sheng, Microglia in Alzheimer ' s disease. *J. Cell Biol.* **217**, 459–472 (2017).
  34. A. Sierksma *et al.*, Novel Alzheimer risk genes determine the microglia response to amyloid- $\beta$  but not to TAU pathology. *EMBO Mol. Med.* **12** (2020), doi:10.15252/emmm.201910606.
  35. T. K. Ulland, M. Colonna, TREM2 — a key player in microglial biology and Alzheimer disease. *Nat. Rev. Neurol.* **14** (2018), pp. 667–675.
  36. S. Parhizkar *et al.*, Loss of TREM2 function increases amyloid seeding but reduces plaque-

- associated ApoE. *Nat. Neurosci.* **22**, 191–204 (2019).
37. K. Schlepckow *et al.*, Enhancing protective microglial activities with a dual function TREM 2 antibody to the stalk region . *EMBO Mol. Med.* **12**, 1–22 (2020).
  38. Y. Zhou *et al.*, Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer’s disease. *Nat. Med.* **26**, 131–142 (2020).
  39. W.-T. Chen *et al.*, Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer’s Disease. *Cell*, 1–16 (2020).
  40. S. J. van der Lee *et al.*, A nonsynonymous mutation in PLCG2 reduces the risk of Alzheimer’s disease, dementia with Lewy bodies and frontotemporal dementia, and increases the likelihood of longevity. *Acta Neuropathol.* **138**, 237–250 (2019).
  41. L. Magno *et al.*, Alzheimer’s disease phospholipase C-gamma-2 (PLCG2) protective variant is a functional hypermorph. *Alzheimer’s Res. Ther.* **11**, 16 (2019).
  42. R. E. Marioni *et al.*, GWAS on family history of Alzheimer’s disease. *Transl. Psychiatry.* **8**, 99 (2018).
  43. S. M. Purcell *et al.*, Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* **460**, 748–752 (2009).
  44. V. Escott-Price, A. J. Myers, M. Huentelman, J. Hardy, Polygenic risk score analysis of pathologically confirmed Alzheimer disease. *Ann. Neurol.* **82**, 311–314 (2017).
  45. V. Escott-Price, A. Myers, M. Huentelman, M. Shoai, J. Hardy, Polygenic Risk Score Analysis of Alzheimer’s Disease in Cases without APOE4 or APOE2 Alleles. *J. Prev. Alzheimer’s Dis.* **6**, 16–19 (2019).
  46. F. Koopmans, P. van Nierop, M. Andres-Alonso, P. D. Thomas, A. B. Smit, SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron.* **103**, 217-234.e4 (2019).
  47. E. Bellou *et al.*, Age-dependent effect of APOE and polygenic component on Alzheimer’s disease. *Neurobiol. Aging* (2020), doi:10.1016/j.neurobiolaging.2020.04.024.

48. G. Leonenko *et al.*, Genetic risk for alzheimer disease is distinct from genetic risk for amyloid deposition. *Ann. Neurol.* **86**, 427–435 (2019).
49. E. A. King, J. Wade Davis, J. F. Degner, Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* **15**, 1–20 (2019).
50. A. Nott *et al.*, Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science (80-. ).* **366**, 1134–1139 (2019).
51. G. Novikova *et al.*, Integration of Alzheimer’s disease genetics and myeloid genomics reveals novel disease risk mechanisms. *bioRxiv*, 694281 (2019).
52. E. Gjoneska *et al.*, Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer’s disease. *Nature.* **518**, 365–369 (2015).
53. J. Park *et al.*, A 3D human triculture system modeling neurodegeneration and neuroinflammation in Alzheimer’s disease. *Nat. Neurosci.* **21**, 941–951 (2018).
54. S. Wray, Modelling neurodegenerative disease using brain organoids. *Semin. Cell Dev. Biol.*, 0–1 (2020).
55. I. Espuny-Camacho *et al.*, Hallmarks of Alzheimer’s Disease in Stem-Cell-Derived Human Neurons Transplanted into Mouse Brain. *Neuron.* **0**, 1066-1081.e8 (2017).
56. J. Hasselmann *et al.*, Development of a Chimeric Model to Study and Manipulate Human Microglia In Vivo. *Neuron.* **103**, 1016-1033.e10 (2019).
57. R. Mancuso *et al.*, Stem-cell-derived human microglia transplanted in mouse brain to study human disease. *Nat. Neurosci.* **22**, 2111–2116 (2019).

## Figure legends

**Fig.1. Risk factors and heritability for Alzheimer's disease (AD).** Whereas 35% of lifetime risk for AD is comprised of modifiable or environmental risk factors, 58-79% of AD risk is genetic. The genetics of AD can be broken down to SNP-based heritability, and other types of genetic variation, including rare variants, structural and copy-number variation, duplications, SNPxSNP interaction, dominance etc.



**Fig.2. Emerging signaling pathways in AD.** TREM2, SORL1 and ABCA7 interact with other genetic risk genes for AD (protein names highlighted in red), impacting microglial function and APP processing. **A) AD pathway in microglia.** TREM2 can bind amyloid- $\beta$  ( $A\beta$ ) that may need to be lipidated by ApoE or ApoJ (*CLU*) and associates with DNAX-activating protein of 12 kDa (DAP12) to constitute intracellular signaling via its immunoreceptor tyrosine-based activating motif (ITAM). The ITAM domain undergoes double phosphorylation by the SRC family kinases (SFK, e.g. LYN) to allow binding of spleen tyrosine kinase (Syk). Syk can phosphorylate phosphoinositide 3-kinase (PI3K) and PLC $\gamma$ 2. Activation of these proteins ultimately leads to calcium and mitogen-activated protein kinase (MAPK) signaling and nuclear factor kappa  $\beta$  (NF- $\kappa\beta$ ) transcription. Protein kinase C (PKC) can also activate proline-rich tyrosine kinase 2 (Pyk2; *PTK2B*), which can activate MAPK signaling, but also associates with Cas scaffold protein family member 4 (CASS4) and focal adhesion kinase (FAK; *PTK2*) to affect actin polymerization, as does Abelson interactor family protein 3 (ABI3). Overall these signaling pathways affect cytoskeletal rearrangements associated with microglial motility and increase phagocytosis. **B) Endocytosis and Alzheimer genes.** *SORL1* can transfer APP to the trans-Golgi network and late endosomes where it undergoes amyloidogenic processing to  $A\beta$ . *SORL1* can also directly bind  $A\beta$  and facilitate its degradation in the lysosomes. Although ABCA7 is involved in cellular lipid homeostasis, e.g. regulating the efflux of lysophosphatidyl choline (LPC) and phosphatidyl choline (PC), ABCA7 can also impact amyloidogenic proteolysis by affecting beta-site APP cleaving enzyme 1 (BACE1) expression levels. Several other AD risk genes involved in endocytic pathway are indicated in red.


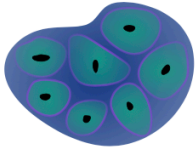



**A: TREM2-related signaling in macrophia**

The diagram illustrates the signaling pathways initiated by TREM2 and CSF1R in macrophia. The cell membrane separates the extracellular space from the intracellular space. Extracellularly, ApoJ and ApoE lipoproteins (orange ovals) bind to TREM2 (red vertical bar) via high-avidity binding. IL34 (orange oval) binds to CSF1R (grey vertical bar). Intracellularly, TREM2 has ITAM domains (red) and DAP12 (pink). CSF1R has an intracellular domain (grey). Signaling molecules include DAG, PKC, NF-κB, FAK, CASS4, Pyk2, MAPK signaling, PLCγ2, Blnk, IP3, Ca<sup>2+</sup> signaling, PI3K, PIP<sub>2</sub>, PIP<sub>3</sub>, ITAM, DAP12, Syk, SFK, and ABI3. The nucleus contains PU.1, NF-κB, and MEF2C transcription factors. The diagram illustrates the pathways leading to phagocytosis, actin polymerization, and cytokine/chemokine production.

**B: AD genes involved in endocytosis**

This diagram illustrates the pathways for the endocytosis and degradation of Amyloid Precursor Protein (APP) and Aβ. APP is synthesized in the Trans-golgi network and can follow two main routes: 1) Secretory pathway: APP is cleaved by α-secretase in the Golgi, releasing soluble APPα. 2) Endocytic pathway: APP is internalized via endocytosis. SORL1 is a key receptor involved in APP trafficking from the Golgi to endosomes. In endosomes, APP can be cleaved by γ-secretase, releasing Aβ. Aβ can then be cleared by ABCA7 (leading to PC and LPC) or taken up by ApoE. Alternatively, Aβ can be degraded in lysosomes by cathepsins (CTD2AP, CTN1B, CTN3). The diagram also shows the role of PICALM and Clathrin in endocytosis, and the involvement of BACE1 in APP processing in late endosomes.

**Fig.3. The opportunities and limitations of commonly used models in AD research.**

	Opportunities	Limitations
<p>2D iPSC cultures</p> 	<ul style="list-style-type: none"> <li>- Human/patient cells</li> <li>- Human A<math>\beta</math> and TAU</li> <li>- Gene editing</li> <li>- Stimulus-response</li> <li>- Screening platform</li> </ul>	<ul style="list-style-type: none"> <li>- Uncontrolled genetic background</li> <li>- Reprogramming &gt; de novo mutations</li> <li>- Limited cell-cell interaction</li> <li>- Transcriptome affected</li> <li>- No full plaques or tangles</li> </ul>
<p>3D models &amp; brain organoids</p> 	<ul style="list-style-type: none"> <li>- Cell-cell interactions</li> <li>- Cortical layers</li> <li>- Circuitry/Electrophysiology</li> <li>- A<math>\beta</math> plaque &amp; tangle formation</li> </ul>	<ul style="list-style-type: none"> <li>- Highly variable</li> <li>- Immature/prenatal cell states</li> <li>- Limited microglia</li> <li>- Limited vascularization</li> <li>- Necrotic core</li> </ul>
<p>Transgenic mice</p> 	<ul style="list-style-type: none"> <li>- Full organism</li> <li>- Mature/aging brain</li> <li>- Controlled genetic background</li> <li>- Gene editing</li> </ul>	<ul style="list-style-type: none"> <li>- Overexpression artifacts</li> <li>- Unwanted genetic alterations</li> <li>- Lengthy experiments</li> <li>- Ethical approval &amp; considerations</li> </ul>
<p>Knock-in mice</p> 	<ul style="list-style-type: none"> <li>- Endogenous expression levels</li> <li>- No overexpression artifacts</li> </ul>	<ul style="list-style-type: none"> <li>- Mouse <math>\neq</math> human</li> <li>- Lack of human-specific interactors</li> <li>- Ethical approval &amp; considerations</li> <li>- Lengthy experiments</li> </ul>
<p>Xenografted mice</p> 	<ul style="list-style-type: none"> <li>- Human cells in 'physiological' context</li> <li>- Complex cell-cell interactions</li> <li>- Cells retain human identity</li> <li>- 1 year+ follow-up</li> <li>- Exposure to relevant pathologies</li> </ul>	<ul style="list-style-type: none"> <li>- Immune compromised background</li> <li>- Human-mouse cell interactions</li> <li>- Human neurons relatively immature</li> <li>- Ethical approval &amp; considerations</li> <li>- Lengthy experiments</li> </ul>



## Box 1 – Glossary

- **Heritability:** the proportion of phenotypic variance that is due to genetic factors.
- **Missing heritability:** the difference between the genetic heritability observed in families and the estimated heritability of identified genetic variants in the population.
- **Core gene:** a mutation in this gene will directly impact disease
- **Peripheral gene:** a mutation in this gene will only indirectly impact disease, most likely through a *trans*-regulatory effect on core genes
- **Core disease pathway genes:** genes directly impacting pathways that determine disease onset
- **Master regulatory gene:** a peripheral gene that regulates the expression or function of several core genes in the disease. Examples include transcription factors, regulatory RNAs or enzymes, or chromatin modifiers.
- **Genotype-phenotype dose-response:** several alleles of a gene impact disease risk, possibly to different degrees, e.g. common and rare, or loss- and gain-of-function variants. Either multiple alleles can affect the same gene or causal alleles are present in different genes that cooperate within the same disease pathway. An example is the amyloid- $\beta$  pathway, where mutations in APP, the presenilins and  $\alpha$ -secretase impact the same pathway and both protective and risk variants have been identified.
- **Polygenic risk score:** a single genetic score indicating a person's risk of developing a trait. Calculated by summing the number of risk alleles present and multiplying this by their effect size, i.e. the weight of disease risk.
- **Linkage disequilibrium:** the observation that specific alleles at a particular genomic locus or region are more often co-inherited within the population than expected by chance.

## Acknowledgements

**Funding:** The authors wish to thank the Dementia Research Institute (UKDRI supported by Medical Research Council (MRC), Alzheimer's Research UK and Alzheimer's Society), MRC Centre for Neuropsychiatric Genetics and Genomics (MR/L010305/1), Welsh Government, Joint Programming for Neurodegeneration (JPND), VIB and KU Leuven (Methusalem grant), the European Union (grant no. ERC-834682 CELLPHASE\_AD), the 'Fonds voor Wetenschappelijk Onderzoek', the 'Geneeskundige Stichting Koningin Elisabeth', Opening the Future campaign of the Leuven Universitair Fonds, the Belgian Alzheimer Research Foundation and the Alzheimer's Association USA. B.D.S. is holder of the Bax-Vanluffelen Chair for Alzheimer's disease.

**Author contributions:** A.S, V.E-P and B.D.S wrote the manuscript.

**Competing interests:** B.D.S. is ad hoc consultant for various companies but has no direct financial interest in the current manuscript.